The CAAPN Network:

$$X = [x_1, x_2, \ldots, x_n]$$

$$PW = W^{previous}$$

$$PB = B^{previous}$$

$$GX = [X, PW, PB]$$

$$GWW = [gww_1, gww_2, \ldots]$$

$$GWB = [gwb_1, gwb_2, \ldots]$$

$$GBW = [gbw_1, gbw_2, \ldots]$$

$$GBB = [gbb_1, gbb_2, \ldots]$$

$$j = (k \cdot l^{max} + l)$$

$$Y_l = AF\left(\sum_{k=0}^{q}\left(X_k \cdot AF\left(\sum_{i=0}^{m}(GX_i GWW_{ij}) + GWB_j\right)\right) + AF\left(\sum_{i=0}^{p}(GX_i GBW_{il}) + GBB_l\right)\right) = AF(Z_l)$$

$$\begin{aligned} Z_l &= \sum_{k=0}^{q}\left(X_k \cdot AF\left(\sum_{i=0}^{m}(GX_i GWW_{ij}) + GWB_j\right)\right) + AF\left(\sum_{i=0}^{p}(GX_i GBW_{il}) + GBB_l\right) \\ &= \sum_{k=0}^{q}(X_k \cdot W_{kl}) + B_l \end{aligned}$$

$$W_{lk} = AF\left(\sum_{i=0}^{m}(GX_i GWW_{ij}) + GWB_j\right) = AF\left(WZ_{lk}\right)$$

$$WZ_{lk} = \sum_{i=0}^{m}(GX_i GWW_{ij}) + GWB_j$$

$$B_l = AF\left(\sum_{i=0}^{p}(GX_i GBW_{il}) + GBB_l\right) = AF\left(BZ_l\right)$$

$$BZ_l = \sum_{i=0}^{p}(GX_i GBW_{il}) + GBB_l$$

Derivatives:

$$\frac{\partial Y_l}{\partial Z_l} = AF'(Z_l)$$

$$\frac{\partial Z_l}{\partial W_{lk}} = X_k$$

$$\frac{\partial Z_l}{\partial B_l} = 1$$

$$\frac{\partial Z_l}{\partial X_k} = W_{kl}$$

$$\frac{\partial W_{lk}}{\partial WZ_{lk}} = AF'(WZ_{lk})$$

$$\frac{\partial B_l}{\partial BZ_l} = AF'(BZ_l)$$

$$\frac{\partial WZ_{lk}}{\partial GWW_{ij}} = GX_i$$

$$\frac{\partial WZ_{lk}}{\partial GX_i} = GWW_{ij}$$

$$\frac{\partial WZ_{lk}}{\partial GWB_j} = 1$$

$$\frac{\partial BZ_l}{\partial GBW_{il}} = GX_i$$

$$\frac{\partial BZ_l}{\partial GX_i} = GBW_{il}$$

$$\frac{\partial BZ_l}{\partial GBB_l} = 1$$

$$\frac{\partial Z_l}{\partial GX_i} = \sum_{k=0}^{q} \left(X_k \cdot \frac{\partial W_{lk}}{\partial WZ_{lk}} \cdot \frac{\partial WZ_{lk}}{\partial GX_i} \right) + \frac{\partial B_l}{\partial BZ_l} \cdot \frac{\partial BZ_l}{\partial GX_i}$$

$$\frac{\partial Z_l}{\partial GX_i} = \sum_{k=0}^{q} (X_k \cdot W_{kl} \cdot AF'(WZ_{lk}) \cdot GWW_{ij}) + AF'(BZ_l) \cdot GBW_{il}$$

$$\frac{\partial Y_l}{\partial GX_i} = \frac{\partial Y_l}{\partial Z_l} \cdot \frac{\partial Z_l}{\partial GX_i}$$

$$\frac{\partial Y_l}{\partial GX_i} = AF'(Z_l) \cdot \left(\sum_{k=0}^{q} (X_k \cdot W_{kl} \cdot AF'(WZ_{lk}) \cdot GWW_{ij}) + AF'(BZ_l) \cdot GBW_{il} \right)$$

$$\frac{\partial Y_l}{\partial GWW_{ij}} = \frac{\partial Y_l}{\partial Z_l} \cdot \frac{\partial Z_l}{\partial W_{lk}} \cdot \frac{\partial W_{lk}}{\partial WZ_{lk}} \cdot \frac{\partial WZ_{lk}}{\partial GX_i}$$

$$\frac{\partial Y_l}{\partial GWW_{ij}} = AF'(Z_l) \cdot (X_k \cdot W_{kl} \cdot AF'(WZ_{lk}) \cdot GWW_{ij})$$

$$\frac{\partial Y_l}{\partial GWB_j} = \frac{\partial Y_l}{\partial Z_l} \cdot \frac{\partial Z_l}{\partial W_{lk}} \cdot \frac{\partial W_{lk}}{\partial GWB_j}$$

$$\frac{\partial Y_l}{\partial GWB_j} = AF'(Z_l) \cdot (X_k \cdot W_{kl} \cdot AF'(WZ_{lk}))$$

The preffered activation function is a trainable bezier spline this is because we are trying to maximize overfitting and the bezier spline will allow us to do this. The bezier spline is defined as follows:

the binomial coefficient is defined as follows:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$B(t) = \sum_{i=0}^{n} \binom{n}{i} (1-t)^{n-i} t^i P_i$$

$$\frac{\partial B(t)}{\partial P_i} = \binom{n}{i} (1-t)^{n-i} t^i$$

$$\frac{\partial B(t)}{\partial t} = \sum_{i=0}^{n} \binom{n}{i} \left(-n(1-t)^{n-1} t^i P_i + (1-t)^{n-i} i t^{i-1} P_i\right)$$

We use $\frac{\partial B(t)}{\partial t}$ as the activation derivative t being the activation input and we use $\frac{\partial B(t)}{\partial P_i}$ to train the points of the spline and allow for it to learn the optimal activation function for the network.